# Wetlands mapping in the Central African Congo River Basin using remote sensing multisource data

## I. Introduction

Wetlands are areas inundated or saturated by surface or ground water at a frequency and duration sufficient to support a prevalence of vegetation adapted for life in saturated conditions (Federal register, Circular 33; 1984). There is no clear boundary between open water and wetland on the one hand and wetland and terrestrial communities on the other hand (Bradbury and Grace, 1983). Wetlands are seen as diverse mosaics of landforms, communities and environments acting as interfaces between terrestrial and aquatic systems (Naiman et al, 1998; Junk and Piedade, 2005). Clement (1905) and Hansen and Di Castri (1992) use the concept of ecotone to represent them as tension zones between plant communities. Junk (1980) and Odum (1985) argue that wetlands have a status of specific ecosystems because soil water saturation, anaerobic conditions, and flooding, provide specific environmental conditions that result in specific biogeochemical processes, organisms specific adaptations, and particularities in community structure and development (Junk and Piedade, 2005; Junk, 1980; Odum, 1981). Flooded forests, also referred to as forested wetlands, are wetlands with a significant component of woody vegetation, living in temporarily or permanently waterlogged or inundated soils (Lugo, 1990; Beard, 1944; Naiman et al., 1998; Trochain, 1957).

The role of tropical wetlands in global climate system, ecosystem regulation, biodiversity conservation, water supply, water quality preservation and food supply has long been recognized (Naiman et al, 1998; Mathews and Fung, 1987; Segers, 1998). This has served as an important justification for promoting international wetland protection frameworks such as the Ramsar Convention (Davis, 1994) and other initiatives under the frame work of the Convention for Biological Diversity (Balmford et al., 2005). However, despite their economical and ecological importance, there are many uncertainties on their extent, distribution, and ecological and physical functions (Mathews and Fung, 1987; Junk and Piedade, 2005; Wang et al 1995). Political instabilities in most humid tropical countries during the last five decades, poor infrastructures, isolation and difficult access may account in part for the scientific inattention (Campbell, 2005).

Although persistent cloud coverage often prevent data acquisition from passive remote sensing systems in equatorial areas, optical remote sensing data have widely been used for deforestation mapping, that is, to separate degraded from intact forest (Achard et al 2001, Hansen and Defries 2004, Hansen et al. 2003). Tropical wetlands however have had little attention from the remote sensing community especially in Central Africa (De Grandi et al., 2001). A number of studies have focused on mapping tropical flooded forests, most of them using active remote sensing data. Wang et al (1995) used radar backscattering for mapping the flooded and non flooded Amazonian forest. Van der Saden and Hoekman (1999) used airborne radar data to support the assessment of tropical primary forests. Hoekman and Quinones (2002) used airborne polarimetric SAR data for biophysical forest type characterization in the Colombian Amazon. De Grandi et al (2000) used wide area radar mosaics for flooded forest mapping at regional scale in the Congo Basin. Hess et al. (2003) used mosaicked L-band synthetic aperture radar (SAR) imagery for dual-season mapping of wetland inundation and vegetation in the Central Amazon. Although these studies have demonstrated the great potential for using radar data for mapping tropical wetlands, many limitations remain to the application of standalone SAR data (Twonsend, 2002) because most of the analyses reported in the literature employed data from airborne SAR missions which are not widely available and a very few works have addressed the need for specifically mapping wetlands in the Congo basin.

Over the years, a growing number of works have been suggesting approaches employing both passive and active remote sensing for better characterization of vegetation (Le Hegart-Mascle et al. 2003; Amarsaikhan and Douglas 2004). In fact, it has been shown that the use of multi-source remote sensing data offers improved classification accuracy as compared to the accuracy achieved by a single source classification (Briem, Benediktsson and Sveinsson 2002; Amarsaikhan and Douglas 2004). However the selection of appropriate analysis methods is still a challenge since conventional parametric statistical

pattern recognition methods are not appropriate in classification of multi-source data (Briem, Benediktsson and Sveinsson 2002).

The availability of optical and radar data as well as digital elevation models derived from passive and/or active remote sensing systems represent an excellent opportunity for mapping wetlands in the tropics as shown by Hamilton et al. (2007) for the Madre de Dios River in Peru. Our work presents result of classification of wetlands in the Congo River basin using multisource remote sensing data including Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+), the Japan Earth Resources Satellite (JERS-1) low and high water radar data and 3 arc seconds elevation data derived from the Shuttle Radar Topography Mission (SRTM).

## 2. Materials and methods

### 2.1 Study site

The Congo Basin extends to approximately 3,000,000 $km^2$ stretching from northern Cameroun to northern Angola. It holds the world's second largest tropical forest with relatively uniform forest over approximately 2,000,000 $km^2$ (Laporte et al, 1998) in six central African countries: Cameroun, Central African Republic, Congo, Equatorial Guinea, Gabon, and the Democratic Republic of the Congo. This research cover a rectangular area comprised between the parallels 5 degrees North and 6 degrees south and between the meridians 13 and 26.5 degrees East inside of the basin (fig.1).

The climate of the region is generally warm and humid with two seasons: a long rainy season and a dry season. The average rainfall is about 1800 mm per year in 115 days (Bultot, 1974). As the basin encompasses the Equator, there is little seasonal variation for regions located 1 degree north or south of the Equator and an inverse seasonal variation far from that line in either side. The northern part of the basin experiences a rainy season from the first half of March to early November while in the south; the rainy season last from September to early June. There is little variation in the temperature throughout the year. The mean temperature during the year is approximately 25 degrees Celsius, but the difference between the high temperature of the warmest month (March) and the coldest month (July) is only 2 degrees Celsius (Bultot, 1974).

The climax ecosystem is a tropical evergreen forest showing little or no seasonal variation (Mayaux et al, 2000; Lebrun and Gilbert, 1954). The highly heterogeneous upper canopy layer is composed of evergreen and shade tolerant species that can reach 35 to 45 meters height. They have irregular and very dense (70-100%) crown coverage which preclude the development of shrubs and grass in the under storey (Mayaux et al, 2000). However, the lower layers have two individual strata composed mainly of regenerating upper canopy layer species (Lebrun and Gilbert, 1954).

At the border of the central basin, semi-deciduous forests become the dominant climax vegetation (Mayaux et al, 2000; Devred, 1958) in mixture with evergreen species in the upper canopy layer (Mayaux et al., 2000). The crown coverage totally closed during the rainy season can become slightly open during the dry season when some species loose their leaves. This favors the development of shade tolerant species in the lower canopy layers (Mayaux et al., 2000).

Secondary forests are complexes of regrowth, fallow and crops with dense undergrowths and regular crown cover. The upper canopy layer is continuous and homogeneous with often *Musanga cecropioides* as dominant specie (Mayaux et al., 2000).

Swamp forests are located along major rivers. They are temporally or permanently inundated and characterized by soils with poor drainage (Mayaux et al, 2000). These forests cover large areas in the center of the study site mainly in the western part of the Democratic Republic of Congo and the northeastern part of the Republic of Congo (Mayaux et al, 2000).

## 2.2. Data sets

### 2.2.1. Landsat Data

Landsat ETM+ data were obtained from the University of Maryland Global land Cover Facility website (http://glcf.umiacs.umd.edu) for two different time periods (1990's and 2000's). We used NASA's Orthorectified global Landsat dataset (Geocover) because they have less than 50-m root mean square location error (Neigh, Tucker and Townsend, 2007 in press). The Geocover coordinate system is defined in Universal Traverse Mercator (Hansen et al, 2008, in press). Additional data were coregistered and resampled to fit the Geocover coordinate system using a ground control point automatic matching algorithm and by bilinear sampling (Hansen et al, 2008, in press).

These Landsat data were normalized using the dark object subtraction (DOS) method to reduce the impact of sensor calibration change, difference in scene illumination and observation angle, variation in atmospheric effects, and phenological variations. A previous 250-m MODIS Vegetation Continuous Field (VCF) Forest / Non forest map (Hansen et al., 2003) was used to perform this normalization for infrared ETM+ bands 4, 5, 7 and for the thermal band 6. A cloud and cloud shadow classification tree model was developed and applied to each normalized Landsat scene to detect and to classify clouds and shadow into low, medium and high confidence categories. The resulting normalized Landsat scenes had reduced radiometric differences due to atmospheric and sun-sensor-target geometric variation (Hansen et al., 2008 in press).

### 2.2.2. Radar data

A mosaic of wide area multi-resolution and multi-temporal Synthetic Aperture Radar (SAR) of JERS-1 over the Central African region was compiled under the frame work of the Global Rain Forest Mapping (GRFM) project (Rosenqvist and Birket, 2002; De Grandi et al., 2000; Mayaux et al., 2002), an international collaborative effort managed by the National Space Development Agency of Japan (NASDA). These include 100-m spacing bi-temporal data resampled from 12.5-m resolution JERS-1 L-band horizontally co-polarized SAR data. These data were acquired in 1996 during January to March (low water) and October to November (high water) periods (De Grandi et al., 2000). They comprised georeferenced and calibrated wide area mosaic of both textural and radiometric information (Rosenqvist and Birket, 2002; Mayaux et al., 2002; De Grandi et al., 2000). For the present work, 8 bit binary image files representing brightness of radar reflections were converted to normalized radar cross section ($\sigma^o$) using the following equation provided by the GRFM project team:

$$\sigma^0 = 20\log_{10}(6DN + 250) + F \qquad (1)$$

- where DN is the digital number representing the brightness of each pixel and F is the calibration factor (-68.2 dB for the present data).

Since the GRF mosaic was provided in direct Mercator projection (datum WGS84), it was reprojected using nearest neighbor resampling method to fit the Geocover data coordinate system.

### 2.2.3. Elevation data

Tropical wetlands are often arranged in linear pattern corresponding to valley bottoms (Saalovara et al., 2005). Topography is therefore important for wetlands characterization (Wolock and McCabe, 1999). A number of metrics that quantify the effects of topography on hydrological processes have been widely considered for hydrological modeling (Wolock and McCabe, 1999). Most of these topographic attributes can be derived from digital elevation data and are classified as primary and secondary elevation derivatives (Moore et al, 1991). Slope gradient (slope) and orientation (aspect) are primary attributes derived from digital elevation data while other attributes such as curvatures are secondary layers derived from these primary attributes.

We used 3 arc seconds (approximately 92 meters) Digital elevation Model (DEM) derived from the NASA Shuttle Radar Topography Mission (SRTM), which was flown onboard of the Space Shuttle mission STS-99 in orbit from February 11 to February 22, 2000 (Hennig et al., 2001; Hamilton et al., 2007). The DEM was derived from interferometric processing of single-pass data collected by a C-band (5.6 cm) SAR (Hamilton et al., 2007; Hennig et al., 2001) with an absolute vertical error less than 16 meters and a relative error less than 10 meters (Brown, Sarabandi and Pierce, 2005). It was reprojected from a spherical coordinate system (latitude and longitude) into Geocover coordinate system (UTM zone 33) and was used to derive a number of metrics by fitting a quadratic surface to the digital elevation data for a 3X3 kernel size window and taking the appropriate derivative using the ENVI software topographical modeling package. These include: slope, aspect, shaded relief, curvatures and Root Mean Square Error.

The slope of a surface represent a change in elevation or the magnitude of elevation gradient in a specific direction, usually, the direction of the steepest path up or down the surface (Rana, 2006).
For an analytical surface:

$$Z = \mathrm{F(x,y)} = ax^2 + by^2 + cxy + dx + ey + f \qquad (2)$$

Where a, b, c, d and f are constants, the slope (S) of the surface is defined as the magnitude of the first derivative of the surface function (De Smith, Goodchild and Longley, 2007):

$$S = \sqrt{\left(\frac{\partial Z}{\partial x}\right)^2 + \left(\frac{\partial Z}{\partial y}\right)^2} \qquad (3)$$

Aspect is defined as the directional component of the gradient vector and is the direction of maximum gradient of the surface at a given point. It is the polar angle described by the two orthogonal partial derivatives of the surface function (Wood, 1996; De Smith, Goodchild and Longley, 2007):

$$\frac{\partial Z}{\partial x} \text{ and } \frac{\partial Z}{\partial y} \qquad (4)$$

Curvatures are single measures of the second order derivatives for an intersecting plane (Wood, 1996). This measure yields a different profile depending on the orientation of the intersecting plane (Wood, 1996; De Smith, Goodchild and Longley, 2007): profile convexity, plan convexity, longitudinal, cross-sectional, maximum and minimum curvatures.

The profile convexity intersects with the plane on the elevation axis and aspect direction and measures the rate of change of the slope along the profile (Wood, 1996). The plan convexity intersects with the x,y plane and measures the rate of change of the aspect along the plane (De Smith, Goodchild and Longley, 2007). These two surface curvature measures are in orthogonal directions; the profile convexity is oriented in the direction of maximum gravity effects and the plan convexity is in the direction of minimum gravity (Wood, 1996).

The longitudinal curvature (intersecting with the plane of the slope normal and aspect direction) and the cross-sectional curvature (intersecting with the plane of the slope normal and perpendicular aspect direction) are measures of the surface curvature orthogonally in the down slope and across slope directions, respectively (wood, 1996; De Smith, Goodchild and Longley, 2007).

The maximum and minimum local curvatures of the surface are derived by intersecting a normal plane with the surface. Evans (1979) as cited by Wood (1996) introduced maximum and minimum curvatures for the above defined analytical surface in any plane as:

$$K_{\min} = -a - b + \sqrt{(a-b)^2 + c} \qquad (5)$$

and $$K_{\max} = -a - b - \sqrt{(a-b)^2 + c} \quad (6)$$

De Grandi et al. (1998) showed that the normalized standard deviation for radar data were a good discriminator for the thematic forest classes and therefore could be used for the characterization of flooded forest because it acts as an edge and point target detector in heterogeneous areas (edge) and will be a measure of the canopy or landform structure in homogeneous areas (De Grandi et al., 1998). Therefore, we integrate in the model, the Root Mean Square Error (RMS) which also indicates how well the quadratic surface fits the actual digital elevation data.

Topographic position is of great importance for wetlands characterization due to location of most floodplain in valley bottom. Therefore, the accuracy of wetlands modeling depends upon the quality of landform extraction and modeling. For the present work, a terrain attribute of elevation with a local meaning was considered. The computation of this derivative requires prior calculation of hydrological topology as defined by flow paths and watershed delineation (MacMillan et al., 2000). We used the D8 method implementing the steepest descent approach (O'Callagan and Mark, 1984; Thompson et al., 1998) to compute the flow direction and the upslope contributing area (flow accumulation). Following O'Callagan and Mark (1984), we extracted channels as all points with accumulated areas above a certain threshold defined by the support area threshold. However, the question of what support area threshold to use is still important because different support areas thresholds return drainage networks of different densities. Tarboton et al.(1991) state that the drainage density (Dd) is inversely proportional to the square root of the support area (Sa) and suggest the use of the smallest possible for which elevation scaling related properties ( the power law, the bifurcation and length law and the area law) still holds. However, they also found that the use of higher support areas thresholds means that slopes are averaged over longer distances and therefore reduces the effects of DEM data errors (Tarboton et al., 1991; Tarboton and Ames, 2001; Tarboton, 1997).

We therefore used six support areas thresholds of 500, 1000, 1500, 10000, 15000 and 20000 grid cells (92 x 92 meters size) in order to obtain maximum stream density layers (with smaller Sa) and to reduce effects of data errors at the same time using higher support areas thresholds. We obtain sets of synthetic stream channels of various densities. A higher support area threshold (20000 grid cells) channel network was used for watershed delineation.

We computed the local topography layer as the measure of the difference in elevation between the closest point along a river or stream and the surrounding points at the hillslope (Williams et al., 2000; Deng, 2007). Each cell was evaluated in term of its absolute difference in elevation and horizontal distance to the nearest channel cell to which it is connected by a flow path (MacMillan et al., 2000). Due to the fact that SRTM was mapping the top of canopy, some anomalies appeared in areas were the flooded forest had a higher tree canopy height and most importantly, degraded forest connected to swamp forests appeared as depressions. Some of these anomalies were however attenuated with the higher support area threshold channel network.

### 2.3. Methodology

A classification tree algorithm (Breiman, Olshen and Stone, 1984) was used to estimate a per-pixel likelihood of wetlands using training of swamp/non swamp to predict wetlands using multisource data layers described in the previous section as independent variables. The classification tree uses a set of hierarchical rules that split data into two groups (child nodes) which are purer then the input group (parent nodes). The algorithm searches for the best univariate split (each decision rule is univariate) by iteratively selecting a threshold on the feature, then, computing the reduction of the residual sum of square, or deviance (Simard et al., 2002). The split that produces the greatest reduction of the deviance is used to divide the data and the process begins for the newly created subsets (Hansen et al., 2003).

The tree is finally generalized (pruning) by cutting branches. A cost-complexity function is used to evaluate each node of the initial decision in term of classification error rate on the training set (Simard et

al., 2002; Breiman, Olshen and Stone, 1984). The weakest nodes are successively cut to produce a sequence of smallest decisions trees which are subsequently tested using the training sets. The tree with an overall smallest misclassification error is selected as final (Simard et al., 2002). Each terminal node is finally labeled with the class which has the maximum proportion of samples (Simard et al., 2002; Breiman et al., 1984). For more details in the decision tree implemented in this study, the reader is referred to Hansen et al. (2003).

Training sets were collected using manual photo-interpretation techniques on the Landsat ETM+ color composition, along a 3 pixels grid lines equally spaced in both X and Y direction to make sure that interclass variability is well sampled. This ensured that spatial variations in homogeneous and heterogeneous areas (where more mixed pixels and discontinuities are present) were both well captured to reduce bias.

Thirty perfectly fit trees were grown independently using 20% of the training samples selected randomly with replacement for each class (bagging) with 0.015% cutoff threshold. This method uses a per pixel voting procedure based on thirty derived classification trees to label outputs. Per nodes likelihoods are used to derive mean class membership likelihood values for each pixel (Hansen et al., 2008, in press). The procedure involves a fitting of a linear regression model to the data in each node in order to generate a mean cover value based on training data present in each node (See Hansen et al., 2003 for more details).

Results are finally ranked using a per pixel comparison with the median result for each pixel. An overall median result is selected from the thirty ranked results to produce the final output map. However, this result layer presented some noises which were particularly obvious in some upland forests. Therefore, a standard deviation map was computed to check noises over the thirty mean maps. Higher values of the standard deviation indicate inconsistency in the reproduction of the results and therefore, a high likelihood of noises. By interactively evaluating this map over the entire study site, a threshold was defined in the standard deviation map in order to eliminate noisy pixels which were replaced by lower ranked results which showed fewer noises across the entire area.

## 3. Results

The thirty bagged classification trees using all the 27 layers explained approximately 54% of the total root node deviance (Table 1). They returned around 670 terminal nodes with a mean classification accuracy of approximately 87% (13.0% misclassification error). All the predictors were used during the classification process each predictor contributing at different level to the reduction of the deviance. The local topography layers were the most important drivers for the classification explaining all together 30.6 % of the total deviance showing the relative importance of topographic position for wetlands occurrence. The first split in the tree used the local topography layer derived using 1000 grids support area threshold which alone explained 22.1% of the total deviance. The Landsat band 5 was used to split the first children nodes and explained 6.8% of the total deviance. The Landsat band 4 was used third and explained approximately 3.4% of the total deviance. The slope layer which contributed for 3.2% to the reduction of the deviance was the fifth important layer. The local topography using the smallest support area threshold contributed for 2.6%.

All the remaining parameters were used for the classification even though their contributions were smaller. However, we didn't create a parsimonious model utilizing a reduced number of parameters because we believed that even though the contribution of some parameters were small, they were targeting some specific areas in the swamp that the five most important parameters were likely to misclassify. Furthermore, it was less likely that the reduction of the number of metrics will provide any substantial advantage to the processing. Table 1 shows the relative contribution of each parameter to the classification.

Table 1: Relative contribution of each parameters in the overall classification process

| Layer Name | Deviance | % Total deviance | % Explained deviance |
|---|---|---|---|
| | | | |
| Local Topography 1000 Grid Support Area threshold | 117427.39 | 22.09940 | 40.92 |
| Landsat Band 5 Time 2 | 36394.73 | 6.84936 | 12.68 |
| Local Topography 20000 Grid Support Area threshold | 21110.37 | 3.97289 | 7.36 |
| Landsat Band 4 Time 2 | 18177.74 | 3.42098 | 6.33 |
| slope | 17082.51 | 3.21487 | 5.95 |
| Local Topography 5000 Grid Support Area threshold | 14262.88 | 2.68422 | 4.97 |
| Landsat Band 5 Time 1 | 12480.10 | 2.34871 | 4.35 |
| Radar high water | 9858.23 | 1.85528 | 3.44 |
| Landsat Band 5 Combined | 6482.99 | 1.22007 | 2.26 |
| Landsat Band 4 Combined | 6282.73 | 1.18239 | 2.19 |
| Local Topography 1500 Grid Support Area threshold | 5055.78 | 0.95148 | 1.76 |
| Local Topography 10000 Grid Support Area threshold | 3739.49 | 0.70376 | 1.30 |
| Landsat Band 4 Time 1 | 2908.42 | 0.54735 | 1.01 |
| Root Mean Square Error | 2668.25 | 0.50215 | 0.93 |
| Maximum Curvature | 2130.16 | 0.40089 | 0.74 |
| Landsat Band 7 Time 2 | 1905.45 | 0.35860 | 0.66 |
| Radar low water | 1676.19 | 0.31545 | 0.58 |
| Minimum Curvature | 1354.68 | 0.25495 | 0.47 |
| Landsat Band 7 Time 1 | 1340.40 | 0.25226 | 0.47 |
| Shaded relief | 1268.53 | 0.23873 | 0.44 |
| Local Topography 15000 Grid Sp A thr | 1177.43 | 0.22159 | 0.41 |
| Aspect | 1009.34 | 0.18995 | 0.35 |
| Landsat Band 7 Combined | 334.48 | 0.06295 | 0.12 |
| Crossectional Convexity | 309.65 | 0.05828 | 0.11 |
| Profile Convexity | 224.20 | 0.04219 | 0.08 |
| Longitudinal Convexity | 177.52 | 0.03341 | 0.06 |
| Plan Convexity | 116.85 | 0.02199 | 0.04 |
| | | | |
| **Total** | **286956.50** | **54.00416** | |

The first output map for wetlands is a per pixel likelihood of wetlands (fig.2) which was generated using the Vegetation Continuous Field (VCF) method as defined by Hansen et al. (2003). It shows the likelihood of non-wetlands over the study site ranging from 0 (absolute confidence of wetlands) to 99 (non wetlands). Overall, results agree well with the visual appearance in the Landsat color composition (fig. 3).

A user defined acceptance threshold was used to separate between wetland and non wetland classes generating a wetland mask for the study site (fig. 4). This threshold was found by interactive evaluation of the model on the entirety of the data. Results show that 44% of the study site (322,356 $Km^2$) is occupied by wetlands which represent the maximum inundable area comprising forested and non-forested wetlands. They predominate particularly in the Lake Tele/Tumba Landscape where they extend to 207,467 $Km^2$accounting approximately for 56% of the total landscape area which represent 373,302 $Km^2$. This vast swamp bloc is located west of the Lake Tumba area across the border between the Democratic Republic of the Congo and the Republic of Congo. Another important swamp forest bloc is located west of the lake Mai-Ndombe in the Democratic Republic of the Congo specifically in the basin of the Lotoi and Lukuru rivers.

## 4. Validation

Field validation for regional scale vegetation maps is still a challenge because of the difficulty to conduct ground truth in such a scale (Mayaux et al, 2002). For the present work, results were compared to existing maps: the Africover map for the Democratic Republic of the Congo (Africover, xxxx) and the vegetation map of Central Africa by Mayaux et al. (2002) which included swamp forest as thematic classes. Contingency matrices and a bivariate maps were computed between the swamp forest mask and the Africover and the GLC 200 vegetation maps for which all the wetlands classes were merged.

The overall accuracy representing agreement in both mapping the wetlands and non wetlands between the Wetlands mask and the Africover vegetation map of the Democratic Republic of the Congo (fig.6) is 71.5% with a user's and producer's accuracy of 47 and 87% respectively and a Kappa coefficient of agreement of 41.3% (Table 2).

Table 2: Vegetation contingency matrix and accuracy results for comparing Africover map and the wetlands mask.

| **Wetlands Mask** | **AFRICOVER** | |
|---|---|---|
| | Wetlands | Non Wetlands |
| | | |
| Wetlands | 128818.96 | 147399.62 |
| Non Wetlands | 19863.90 | 291772.32 |
| **Accuracy** | | |
| **Accuracy** | **Wetlands (%)** | **Non-Wetlands (%)** |
| User's accuracy | 46.64 | 66.44 |
| Producer's accuracy | 86.64 | 93.63 |
| | | 71.5% |
| Kappa coefficient | | 0.41 |

AFRICOVER Semi-deciduous forest class was the most confused with the swamp forest with a total 20.2% disagreement (Table 3). Several factors can explain these disagreements. First, an area of swamp forest in the Mankanza and Bomongo territories in the Democratic Republic of the Congo was probably misclassified by the Africover operators (fig.9). Second, ribbons of swamp forests along medium sized rivers in the Maringa-Lopori-Wamba CARPE landscape were classified as non swamp and semi-deciduous forests (fig. xx).

Table 3: Vegetation contingency matrix and accuracy results for comparing Africover map and the wetlands mask by classes.

| **AFRICOVER Classes** | **Wetlands Mask** | | **% Error** | **% Agreement** |
|---|---|---|---|---|
| | Wetlands | Non-Wetlands | | |
| | | | | |
| Dense Humid | **14509.13** | 18717.29 | 2.5 | 3.2 |
| Semi-deciduous | **118462.02** | 187944.53 | 20.2 | 32.0 |
| Woody Savanna | **7874.16** | 48364.48 | 1.3 | 8.2 |
| Dense Dry | **36.57** | 146.44 | 0.0 | 0.0 |
| Secondary Forest | **6388.99** | 36048.85 | 1.1 | 6.1 |
| Savanna | **128.75** | 550.72 | 0.0 | 0.1 |

| | | | | |
|---|---|---|---|---|
| Wetland | 128818.96 | **19863.90** | -3.4 | 21.9 |
| | | | | |
| | **276218.59** | **311636.21** | **28.5** | **71.5** |

Third, small ribbons of swamp forests along small river networks are not taken into account in the Africover map whereas they are well evidenced in the wetland mask. Fourth, the boundaries between forests types are sometime fuzzy in case of visual interpretation such as in the case of the Africover map. Fifth, the color composition reveals clearly evidence of misclassification for some other swamp forest in the Africover map. Sixth, some discrepancies can result from overestimation or underestimation of wetlands during the thresholding process while creating the wetland mask.

The comparison with the regional Global Land Cover 2000 vegetation map of Central Africa (fig.4) shows an overall correspondence of 62.3% with a users and producer's accuracy of 45.5 and 82.3% respectively and a kappa coefficient of agreement of 0.575. The main source of difference is due to misclassification of swamp forests in the lake Tumba / Lake Tele Landscape where the regional radar map classified large swamp areas as upland dense humid forests. In addition, the radar map was unable to detect swamp forests along medium sized and small rivers (fig.10).

Table 4: Vegetation contingency matrix and accuracy results for comparing the Global Land Cover 2000 map and the wetlands mask.

| **Wetland Mask** | **GLC 2000 Regional map** | |
|---|---|---|
| | Wetlands | Non wetlands |
| | | |
| Wetlands | 96569 | 115698 |
| Non Wetlands | 20789 | 129090 |
| | | |
| | 117358 | 244788 |
| **Accuracy** | | |
| **Accuracy** | **Wetlands (%)** | **Non-Wetlands (%)** |
| User's | 45.49 | 52.74 |
| Producer's | 82.29 | 86.12 |
| Overall | | 62.31% |
| Kappa | | .575 |

The Global Land Cover 2000 vegetation map of Central Africa 2000 compare also very well with Africover map with an overall correspondence of 66%, user's and producer's accuracy of 78 and 74% for non wetlands. However, there is less agreement on mapping wetlands with user's and producer's accuracy of respectively 41 and 47%. The kappa coefficient of agreement is also very low (0.20) showing perhaps that these two maps may probably agree more by chance. Table 5 shows results of accuracy assessment for agreement between the Africover map and the Global land Cover 2000.

Table 5: Vegetation contingency matrix and accuracy results for comparing Africover map and the Global land Cover 2000 vegetation map.

| GLC 2000 Regional Map | AFRICOVER | |
|---|---|---|
| | **Wetlands** | **Non Wetlands** |
| | | |
| **Wetlands** | 27,913 | 31,711 |
| **Non-Wetlands** | 39,941 | 113,566 |
| | | |
| | 67,854 | 145,276 |
| | | |
| **Accuracy** | | |
| | | |
| **Accuracy** | **Wetlands (%)** | **Non Wetlands (%)** |
| Producer | 41.14 | 78.17 |
| User's | 46.81 | 73.98 |
| Overall | | 66.38% |
| Kappa | | 0.1995 |

Comparison with GPS points

| Swamp Mask | Truth (Pixels) | |
|---|---|---|
| | **Non Wetlands** | **Wetlands** |
| **Non wetlands** | 4066 | 96 |
| **Wetlands** | 487 | 1050 |
| | | |
| | **4553** | **1146** |
| **Accuracy** | | |
| | | |
| **Accuracy** | **Wetlands (%)** | **Non Wetlands (%)** |
| Producer | 91.62 | 89.30 |
| User's | 68.31 | 97.69 |
| Overall | | 89.77 |
| Kappa | | 0.7176 |

**Conlusion**

The use of multisource data for wetlands mapping in the Congo basin improved significantly the characterization of wetlands in the Congo basin. The process is fully automated except for training samples collection which is sometime labor intensive and requires the interpreter's knowledge of the region's vegetation and landforms. Although it was reported in the literature that SAR data are best suited for mapping wetlands, the backscattering response from upland forests sometime overlaps those of wetlands making it hard to discriminate between these two classes. The use of multisource data contributed to ameliorate the classification due to the complementary characteristics of multisource data.

The Decision Tree analysis tool showed that, in this study, the DEM derivatives and landsat data were the most important drivers for wetlands characterization in the Central Africa region. This shows the importance of topographical position for wetlands occurrence.

Overall, results agree well with the visual appearance of floodplains in the Landsat color composition. However, validation of the accuracy of wetlands characterization is still a critical issue. For the present study, due to lack of ground information, results were compared with existing maps (the AFRICOVER and the Regional Global Land Cover 2000 Central Africa Wetlands product). Results showed good correspondence with these two products even though there was important disagreement in specific areas due to inability of single sensors data used to generate both products to fully capture the variability of flood plains. Even if the confidence in mapping wetlands seems to be satisfactory with our methodology, field information are still needed to support the accuracy of the classification.

**References**


Achard F., Eva H. & Mayaux P. (2001). Tropical Forest mapping from coarse resolution data: Production and accuracy assessment issues. *International Journal of Remote Sensing*, 22 (14) 2741-2762.

Amarsaikhan D. & Douglas T. (2004). Data fusion and image classification. *International Journal of Remote Sensing* 25 (17), 3529-3539.

Balmford A., Bennun L., Brink B., Cooper D., Côté C.M., Crane P., Dobson A., Dudley N., Dutton I., Green R.E., Gregory R.D., Harrison J., Kennedy E.T., Kremen C., Leader-Williams N., Lovejoy T.E., Mace G., May G., Mayaux P., Morling P., Phillips J., Redford K., Ricketts T.H., Rodríguez J.P., Sanjayan M., Schei P.J., Van Jaarsveld A.S., Walther B.A., 2005. The Convention on Biological Diversity's 2010 Target. *Science*, 307 ( 5707): 212 – 213.

Beard J. S. (1944). Climax vegetation in tropical America, *Ecology* 25 (2), 127-158.

Besag J. & Diggle P.J. (1977). Simple Monte carlo test for spatial pattern. *Applied Statistics* 26 (3), pp. 327-333.

Bradbury I.K. & Grace J. (1983). Primary production in wetlands. In Gore A.J.P. (ed). *Ecosystems of the World: Mire: Swamp, bog, fen and moor* (pp 285-310). Amsterdam: Elsevier.

Briem G. J. Benediktsson J. A., Sveinsson J. R. (2002). Multiple classifiers applied to Multisource Remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing* 40 (10) 2291-2299.

Breiman L., Olshen J.H., Stone C.J. (1984). *Classification and Regression Trees*, ed. Chapman and Hall, New York.

Brown C.G., Sarabandi K. & Pierce L.E. (2005). Validation of the Shuttle Radar Topography Mission height data. *IEEE Transaction in Geoscience and Remote Sensing* 43(8):1707-1715.

Bultot F., (1974): *Atlas climatique du bassin zaïrois. Quatrième partie: pression atmosphérique, vent en surface et en altitude, température et humidité de l'air en altitude, nébulosité et visibilité, propriétés chimiques de l'air et des précipitations et classifications climatiques*, - Publ. I.N.E.A.C., 193 maps, Brussels.

Campbell D. (2005). The Congo River basin. In L.H. Fraser and P. A. Keddy (Eds). *The World largest Wetlands: Ecology and Conservation* (pp. 149-165). Cambridge: Cambridge University Press.

Clement F. E. (1905). *Research method in Ecology*. Lincoln NE: University Publishing Co.

Davis T. J. (1994). *The Ramsar Convention Manual. A guide to the Convention of Wetlands of International Importance especially as waterfowl habitat.* Ramsar Convention Bureau, Gland, Switzerland.

De Grandi, F., Mayaux, P., Rosenqvist, A., Rauste, Y., Saatchi S., Simard, M., Leysen M., (1998). Flooded Forest Mapping at Regional Scale in the Central Africa Congo River Basin: First Thematic Results Derived by ERS-1 and JERS-1 Radar Mosaics. .Retrieval of Bio- and Geo-Physical Parameters from SAR Data for Land Applications Workshop, Noordwijk, 21-23 October 1998.

De Grandi G. F., Mayaux P., Rauste Y., Rosenqvist A., Simard M. & Saatchi S., (2000). The Global Rain Forest Mapping Project JERS-1 radar mosaic of tropical Africa: development and product characterization aspects. *IEEE Transaction on Geosciences and Remote Sensing* 38, 2218-2233.

Deng Y. (2007). New trend in digital terrain analysis: landform definition, representation and classification. *Progress in physical geography* 31(4): 405-419.

De Smith M. J., Goodchild M.F., Longley P.A. (2007). Geospatial Analysis - a comprehensive guide. 2nd ed. Splint, London, UK.

FAO. 2005. Land Cover Classification System (LCCS). Classification Concepts and User Manual. Software version 2.0. A. Di Gregorio, Food and Agriculture Organization of the United Nations (FAO), Rome.

Finlayson, C.M., Davidson, N.C., Spiers, A.G. & Stevenson, N. J. (1999). Global wetlands inventory – status and priorities. *Marine and Freshwater Research* 50, 717-727.

Hamilton S.K., Kellndorfer J., Lehner B., Tobler M. (2007). Remote sensing of floodplain as a surrogate for biodiversity in a tropical river system (Madre de Dios, Peru). *Geomorphology* 89: 23-38.

Hansen A. J. & Di Castri F. (1992). *Landscape Boundaries: consequences for Biotic Diversity and Ecological Flows. Ecological Studies*. Berlin: Spinger Verlag.

Hansen M.C., Roy D.P., Lindquist E., Adusei B., Justice C.O., Alstatt A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin. *Remote Sensing of Environment* (in press).

Hansen M. C. & Defries R. S. (2004). Detecting long term global change using continuous fields of trees-cover maps from 8-km Advanced Very High Resolution Radiometer (AVHRR) data for the year 1982-99. *Ecosystems* 7, 695-716.

Hansen M. C., Defries R. S., Townsend R.R.G. Carroll M., Dimicelli C. & Sohlberg R.A. (2003). Global percent of trees cover at a spatial resolution of 500 meters: First result of the MODIS vegetation continuous fields algorithm. *Earth Interactions* 7, 695-716.

Hansen M. C., DeFries R. S., Townsend J. R. G., Sohlberg, R. A., Dimiceli C., Carrol M. (2002). Toward an operational MODIS continuous field of percent tree cover algorithm: example using AVHRR and MODIS data. *Remote Sensing of Environment* 83, 303-319.

Hennig T.A., Kretsch J.L., Pessagno C.J., Salamonowicz P.H. & Stein W.L. 2001. The shuttle radar topography mission in *Digital Earth Moving, Proceedings of First International Symposium, DEM 2001: Manno, Switzerland, September 5-7, 2001*, Lectures notes in Computer Science, Vol. 2181 / 2001, Springer Berlin, pp. 65-77.

Hoekman D. H. & Quinones M. J. (2002). Biophysical forest type characterization in the Colombian Amazon by airborne polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing* 40 (6) 1288-1300.

Junk W. J. (1980). General Aspects of Floodplain Ecology with Special Reference to Amazonian Floodplains. In. Junk W J. (Ed), *Ecological Studies 126. Ecology of a pulsing system* (pp. 23-46). Berlin: Spinger.

Junk W.J. & Piedade M.T.F. (2005). The Amazon River basin. In L.H. Fraser and P. A. Keddy (Eds), *The World largest Wetlands: Ecology and Conservation* (pp. 63-117) Cambridge: Cambridge University Press.

Laporte N., Goetz S.J., Justice C.O., & Heinicke M. (1998). A new land cover map of Central Africa derived from multiresolution, multitemporal AVHRR data, *International Journal of Remote Sensing*, 19 (18) 3537-3550.

Lebrun J. and Gilbert G. (1954). *Une classification écologique des forêts du Congo*. Série Scientifique Nº 63, INEAC, Bruxelles, 44 pp.

Le Hegart-Mascle S., Richard D. & Ottle C. (2003). Multi-scale data fusion using Dempster-Shafer evidence Theorie. *Integrated Computer Aided Engineering* 10, 9-22.

Lugo, A.E., S. Brown & Brinson M.M. (1990). Synthesis and search for paradigms in wetland ecology. In A.E. Lugo, M. Brinson and S. Brown (eds.), *Ecosystems of the world: Forested wetlands* (pp. 447-460). Amsterdaam: Elsevier.

MacMillan R.A., Pettapiece W.W., Nolan S.C., Goddard T.W., 2000. A generic procedure for automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy sets and Systems* 113:81-109.

Matthews E. & Fung I. (1987). Methane emission from natural wetlands: global distribution, area and environmental characteristics of sources. *Global Biogeochemical Cycles* 1 (1): 61-86.

Mayaux P., De Grandi G., Malingreau J.P. (2000). Central African Forest Cover Revisited: A multisatellite Analysis. *Remote Sensing of Environment* 71, 183-196.

Mayaux P., De Grandi G., RAUSTE Y., Simard M., Saatchi S. (2002). Large-scale vegetation maps derived fro the combined L-band GRFM and C-band CAMP wide area radar mosaics of Central Africa. *International Journal of Remote Sensing*, 23 (7) 1261-1282.

Naiman, R.J., Fetherston K.L., McKay S., & Chen. J. (1998*). Riparian forests*. In R.J. Naiman and R.E. Bilby (Editors), *River Ecology and Management: Lessons from the Pacific Coastal Ecoregion* (pp. 289-323). New York: Springer-Verlag

Neigh C.S.R, Tucker C.J. & Townsend J.R..G (2007). North American vegetation dynamics observed with multi-resolution satellite data. *Remote Sensing of Environment* (in press).

O'Callagan J.F. & Mark D.M (1984). The extraction of drainage networks from digital elevation data. *Computer Vision Graphic Image Processing* 28:328-344.

Odum E.P., 1985. Trends expected in stressed ecosystem. *BioScience*, 35:419-422.

Rana S. (2006). Use of Plan Curvature Variation for the identification of ridges and channels, in Progres in Spatial Data Handling, 12th International Symposium on Spatial Data Handling, A. Riedl, W. Kainz and G.A. Elmes (eds), Springer Berlin Heidelberg, Germany, pp. 790-804.

Rosenqvist A. & Birkket C.M., 2002. Evaluation of JERS-1 SAR mosaics for hydrological applications in the Congo river basin. . *International Journal of Remote Sensing*, 23 (7) 1283-1302.

Salovaara, K., Thessler, S., Malik, R.N. & Tuomisto, H. (2005) Classification of Amazonian primary rain forest vegetation using Landsat ETM+ satellite imagery. *Remote Sensing of Environment* 97: 39-51.

Segers R. (1998). Methane production and methane consumption: A review of processes underlying wetland methane fluxes. *Biogeochemistry* 27: 35-60.

Simard M., De Grandi J.F., Saatchi S. & Mayaux P. (2002). Mapping tropical coastal vegetation using JERS-1 radar data with a decision tree classifier. *International Journal of Remote Sensing* 23 (7): 1461-1474.

Tarboton D.G (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models. *Water Resources Research* 33(2): 300-319.

Tarboton D.G. and Ames D.P. (2001). Advances in mapping of flow networks from digital elevation data. *World water and Environmental Resources Congress, May 24, 2001*, Orlando, Florida, USA.

Townsend P. A. (2002). Estimating forest structure in wetlands using multitemporal SAR. *Remote Sensing of Environment* 79, 288-304.

Trochain (1957). Accord interafricain sur la definition des types de vegetation de l'Afrique Tropicale. *Bulletin de l'Institut des Etudes Centrafricaines Nouvelle Série* 13-14, 55-93.

Van der Sanden J. J. & Hoekman D. H. (1999). Potential airborne radar to support the assessment of land cover in tropical rain forest environment. *Remote Sensing of Environment* 68, 26-40.

Wang Y., Hess L. L., Filoso S. & Melack J. M. (1995). Understanding the radar backscattering from flooded and non flooded forest: results from canopy backscatter modeling. *Remote Sensing of Environment* 54, 324-332.

Williams W.A., Jensen M.E., Winne C.J. & Redmond R.L. (2000). An automated technique for delineating and characterizing valley-bottom settings. *Environmental Monitoring and Assessment* 64:105-114.

Wood, J. (1996). The Geomorphological Characterization of Digital Elevation Models, Ph. D. Thesis, University of Leicester, Department of Geography, Leicester, UK.

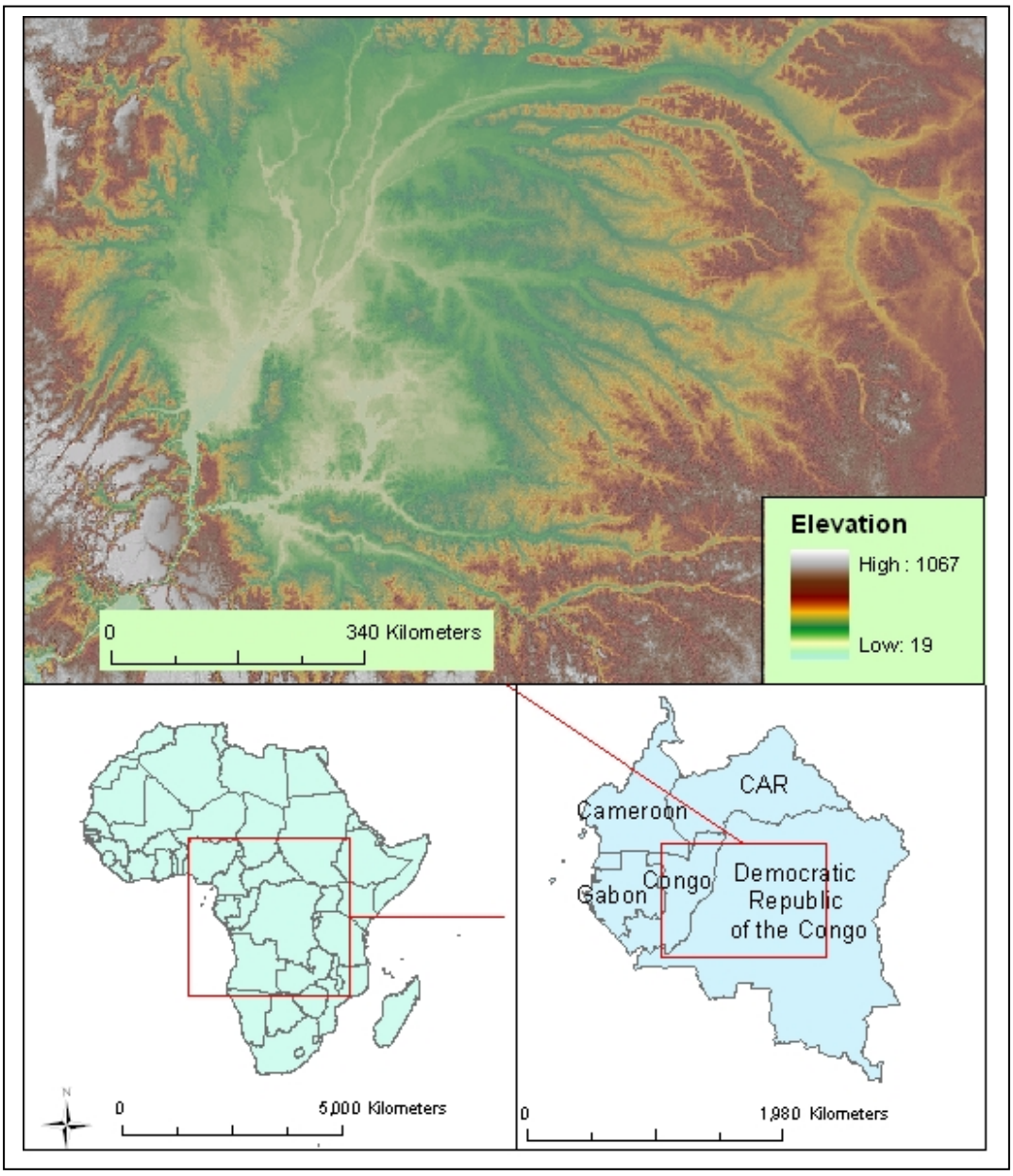


**Figure 1**: Location of the study site

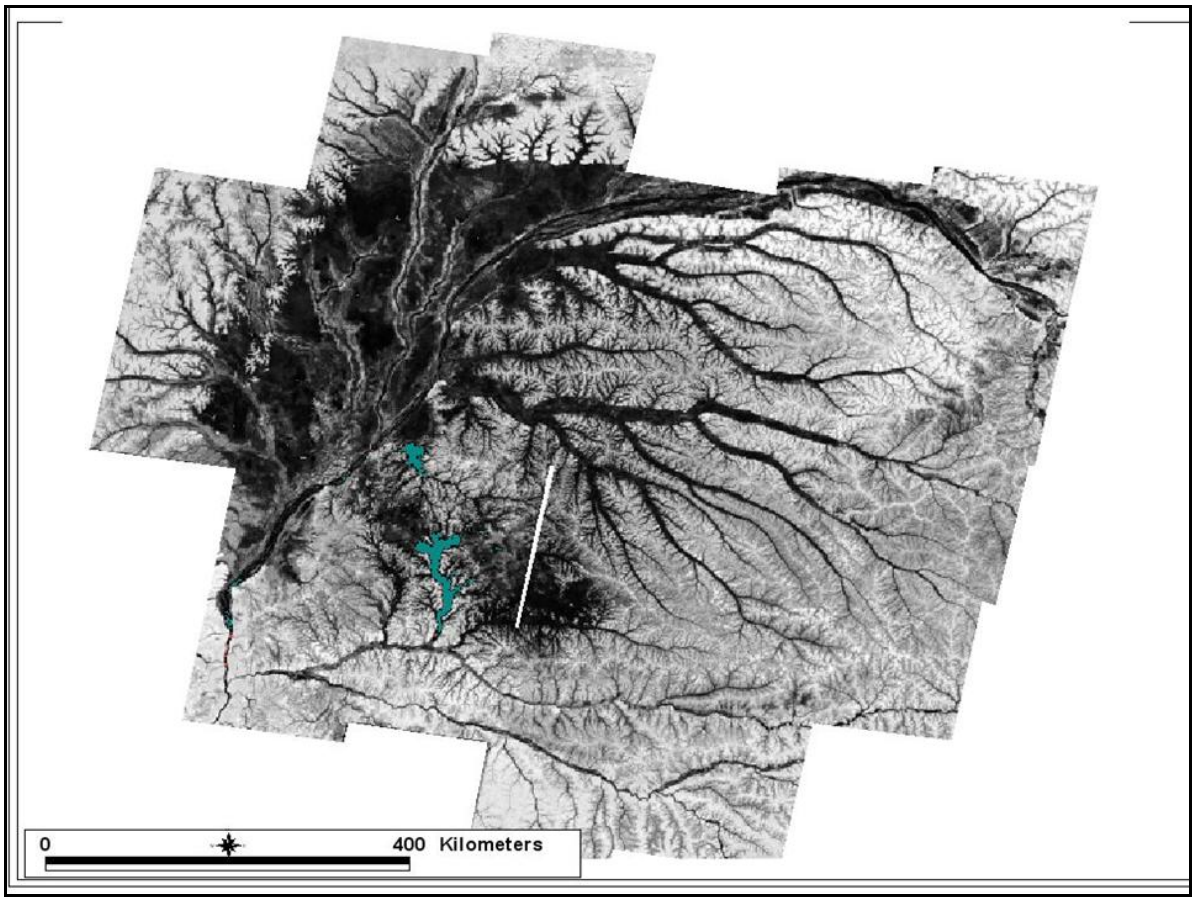


**Figure 2:** Continuous field result showing the Likelihood of Non wetlands for the study site.

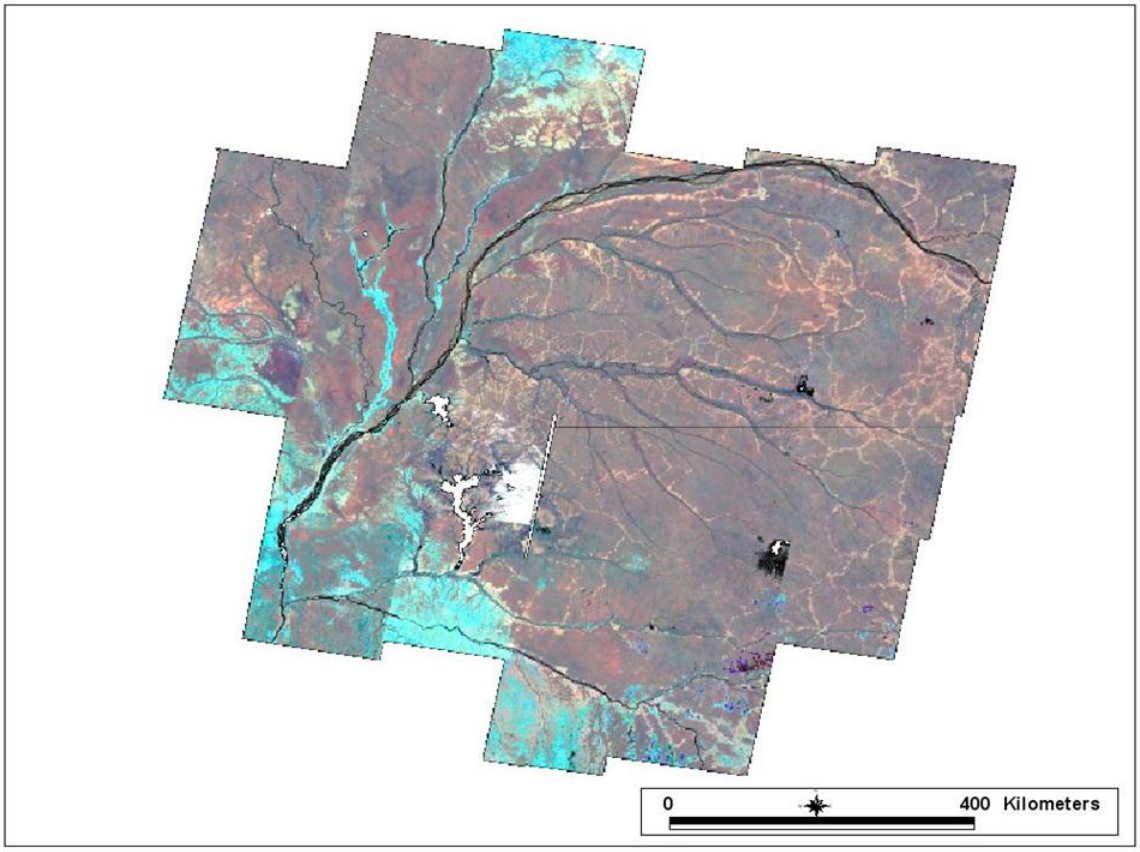


**Fig. 3:** Comparison of the Continuous likelihood map with the color composition

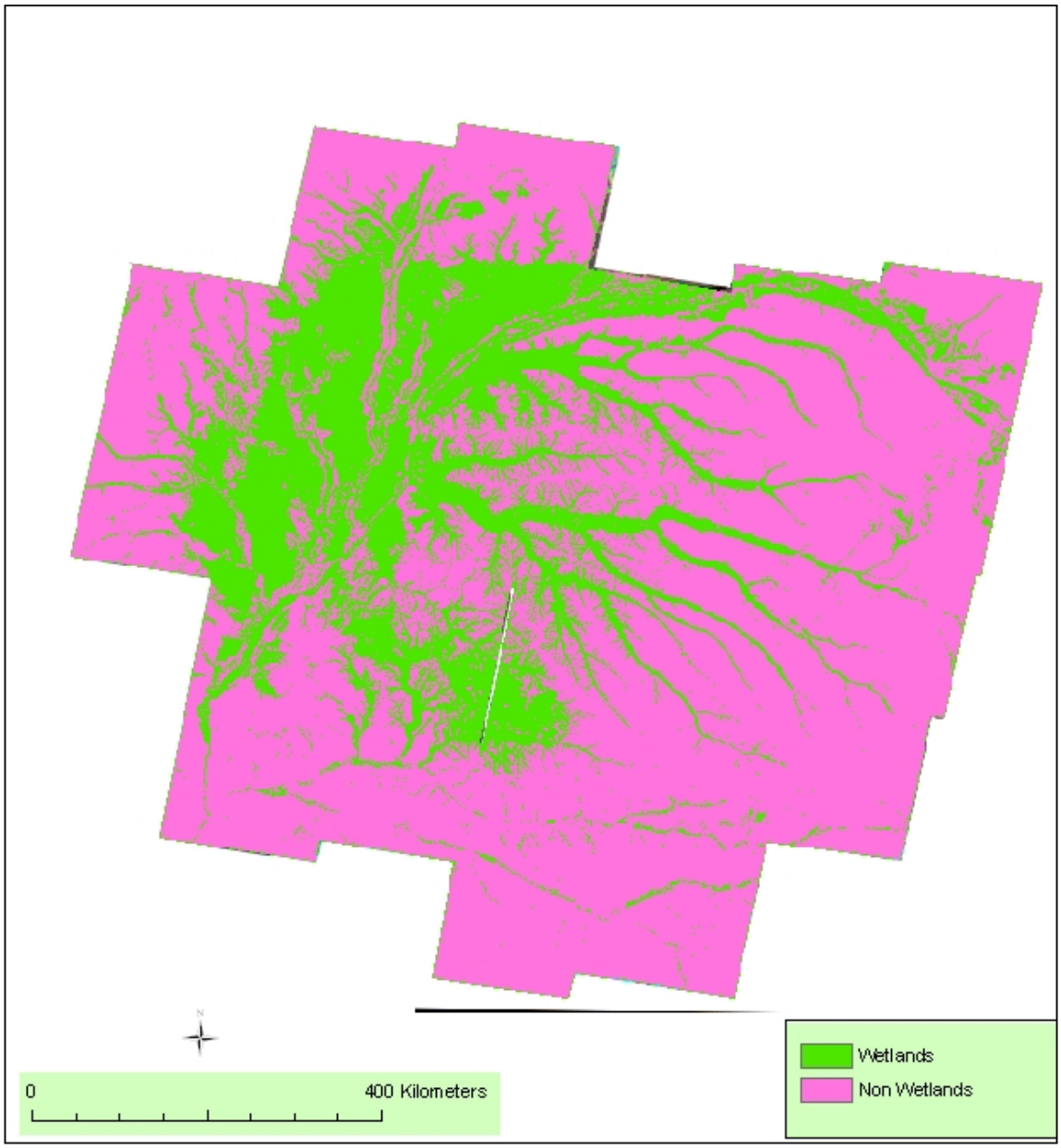


**Figure 4:** Wetlands mask derived by thresholding the Continuous field result to 49%

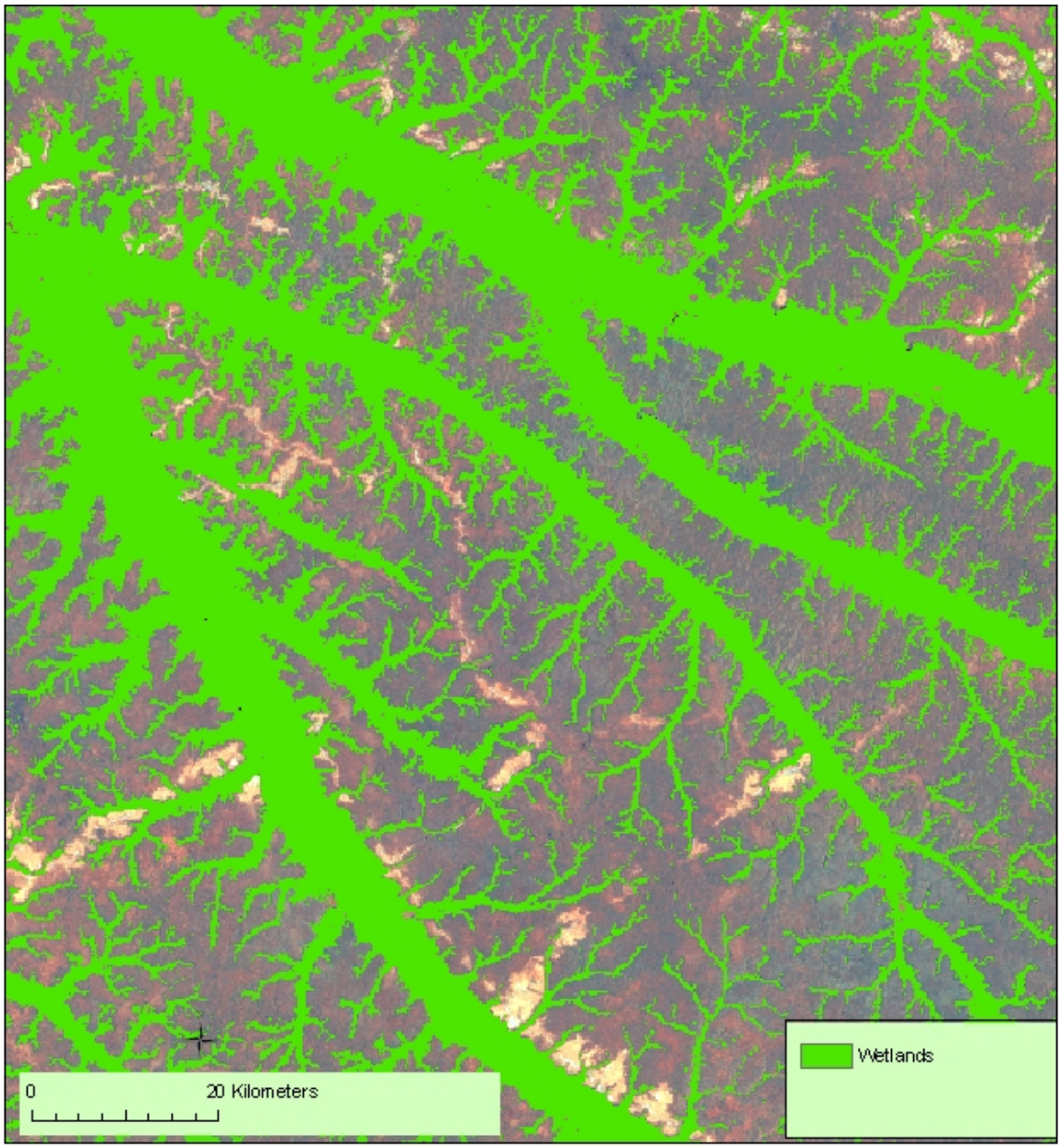


**Figure 5:** A closer view of the wetlands mask

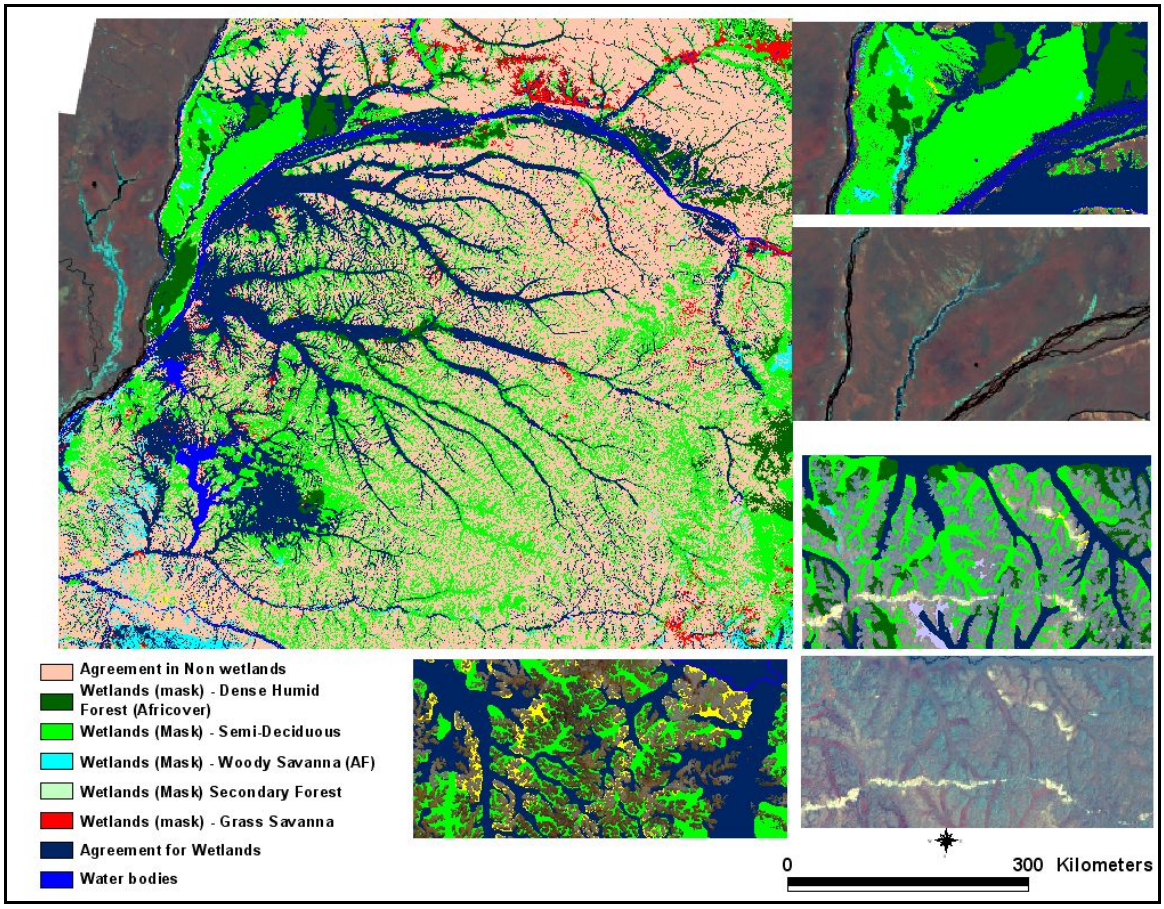


**Figure 6:** Bivariate map for accuracy assessment against the AFRICOVER map

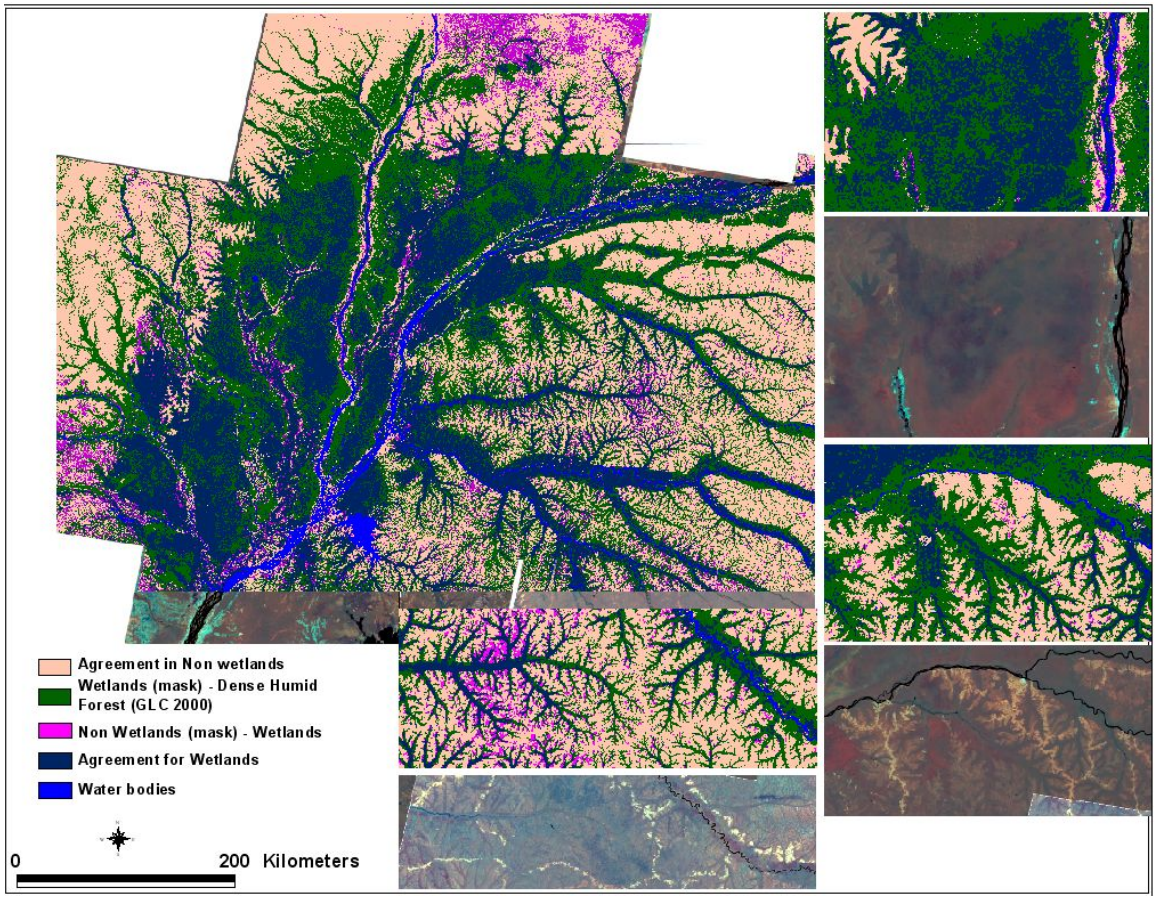


**Figure 7:** Bivariate map for accuracy assessment against the GLC 2000 Wetlands map of Central Africa